Full-chip source and mask optimization

Min-Chun Tsai*, Stephen Hsu, Luoqi Chen, Yen-Wen Lu, Jiangwei Li, Frank Chen, Hong Chen, Jun Tao, Been-Der Chen, Hanying Feng, William Wong, Wei Yuan, Xiaoyang Li, Zhipan Li, Liang Li, Russell Dover, Hua-yu Liu, Jim Koonmen
ASML Brion Technologies, 4211 Burton Dr., Santa Clara, CA, 95054, USA.

ABSTRACT

A cost-effective technique for full-chip source and mask optimization is proposed in this paper. This technique has two components: SMO source optimization for full-chip and flexible mask optimization (FMO). During the technology development stage of source optimization, a novel pattern-selection technique was used to identify critical clips from a full-set of design clips; SMO was then used to optimize the source based on those selected critical-clips. This pattern-selection technique enables reasonable SMO runtime to optimize the source that covers the full range of patterns. During the process development stage and product tapeout stage, FMO is employed. The FMO framework allows the use of different OPC computational techniques on different chip areas that have different sensitivities to process variations. Advanced OPC methods are applied only where they are needed, therefore achieving optimum process performance with the least tapeout and mask cost.

Keywords: Full Chip, SMO, FMO, MB-SRAF, PW OPC, LMC.

1. INTRODUCTION

Below the 32 nm node, manufacturable designs can only be achieved by using an increasing number of advanced computational lithography (CL) techniques required for source and mask optimization especially during the OPC tape-out. These CL techniques typically involve increased computational load and complexity, and as a result, tape-out time and costs increase. This paper examines an approach called full-chip source mask optimization that is designed to enable the targeted and efficient use of these advanced techniques that might otherwise be limited due to high computational and mask costs.

Full-chip source mask optimization is an approach that strives to achieve an optimum solution among the competing requirements of manufacturing process window, mask complexity and cost, and tape-out throughput time. Its two components are source optimization (SMO) for full-chip and flexible mask optimization (FMO). SMO is used to optimize the source given a representative set of clips, and FMO achieves the user-required process window (PW), MEEF and mask complexity with acceptable runtime during the full-chip product tape-out.

SMO has been widely accepted and, combined with the FlexRay programmable illuminator from ASML, has enabled user-tunable off-axis illumination for optimum resolution based on actual designs. Prior publications [1-5] have demonstrated how this is achieved using a set of manually selected design clips. With the number of design clips increasing to over hundreds or even thousands of clips, SMO runtime for all those clips becomes unmanageable. This paper proposes a novel method using a diffraction order technique to select a representative subset of clips from the full set of design clips. This improves SMO runtime greatly without compromising litho performance, and the optimized source is qualified for full-chip designs. Details of this full-chip SMO technique for source optimization are discussed in Section 2.

FMO is a framework employed during the OPC tape-out flow where advanced computational methods for mask correction are applied only where they are most needed, thereby balancing performance and cost. Computationally less expensive methods are used for design areas where a satisfactory process window can be achieved, and more expensive techniques are used for the remaining critical areas where such techniques are the only viable means of achieving the desired process window.

*mitsai@brion.com; phone 1 408 200-3738; fax 1 408 653-1501

Optical Microlithography XXIV, edited by Mircea V. Dusa, Proc. of SPIE Vol. 7973,
79730A · © 2011 SPIE · CCC code: 0277-786X/11/$18 · doi: 10.1117/12.881633

These techniques – like model-based freeform SRAF [6, 7] -- will be discussed in this paper. Essential to this FMO approach is to resolve any boundary conflicts between the different methods, so that no new defects are created on the boundaries. Details of this FMO framework are discussed in Section 3.

With the combinations of SMO for full-chip source optimization and FMO for full-chip mask optimization, this technique of full-chip source mask optimization achieves the required lithography performance with the benefit of acceptable cost. It is a production-worthy technique. Several results based on realistic data are presented.

2. SMO SOURCE OPTIMIZATION FOR FULL-CHIP

2.1 Technique

By co-optimizing the scanner source and mask simultaneously, SMO [1-4] has demonstrated freeform sources and manufacturable wafer exposures of advanced design rules [5]. Those successes were based on SRAM designs with limited patterns. SMO is computational demanding and in practice only limited patterns are used during SMO optimization. In addition, in order to maximize the process window, SMO generated mask can be very complicated. To make SMO applicable for full-chip implementation, two requirements need to be met. One is that the source optimized with selected patterns need to be optimum for the full chip. Another is that the generated source has to achieve the required process window in a full chip application with the least mask complexity.

In this paper, both requirements are met by developing an innovative algorithm for pattern selection and the method of integrating RET/MB-OPC recipes during the SMO source evaluations, as illustrated in Figure 1. It is important to use the pattern selection algorithm to optimize SMO runtime; and to evaluate the SMO source in conjunction with RET/MB-OPC recipes to reduce mask complexity and achieve desired process window for full-chip application. This approach also ensures that the mask optimization technique to be used for actual full-chip manufacturing is compatible and essentially co-optimized with the optimum source.

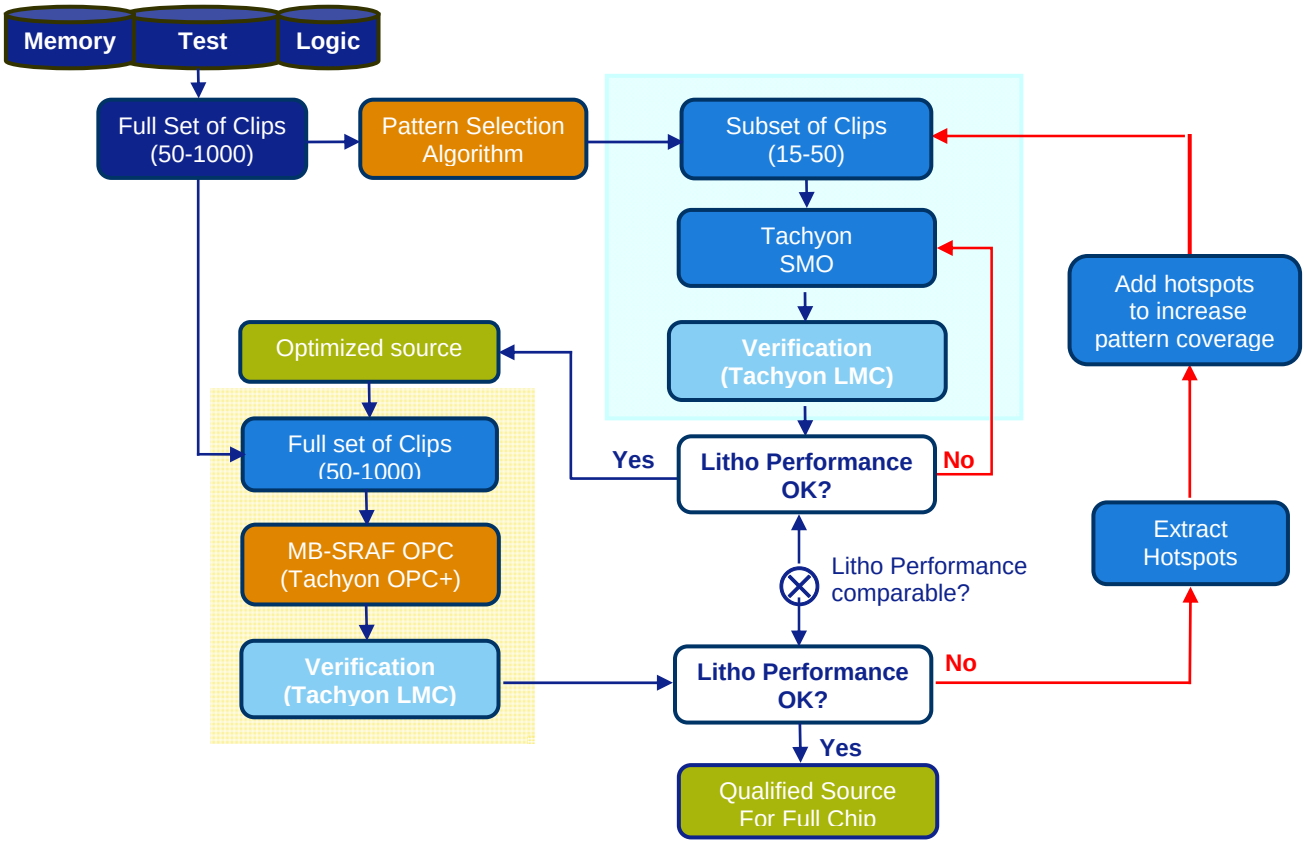

Figure 1. Flow of SMO source optimization for full-chip.

The pattern selection algorithm is based on diffraction orders. The algorithm requires only the design clips patterns as the input, and automatically selects a subset of patterns that best presents all the diffraction orders of interest in the full pattern set. Through this pattern selection, hundreds or thousands of critical design patterns used for process development, e.g., memory cell, test chip and logic patterns, can be reduced to manageable amount of representative patterns for SMO source optimization. The optimized source was then used with the MB-OPC recipe using the full set of clips. The lithography performance result from all the patterns has to be comparable to the SMO lithography performance result from the representative patterns. From the full-pattern set simulation, additional hot-spots maybe added for SMO source optimizations. Result from this method is a qualified SMO source for the full chip application.

2.2 Results

Two test cases were demonstrated here to show the effectiveness of this full-chip SMO source optimization. The first case was a 22 nm technology node logic contact layer .The case has thirty eight (38) pattern clips (SRAM, through pitch test structures and random logic designs) in-all, and after pattern selections, the representative selected-clips were reduced to just 6 clips. Figure 2 shows comparisons of runtime and depth of focus (DOF) results among the full pattern set, the selected-clips and the conventional manually-selected-clips. The conventional manually-selected-clips include a SRAM and the densest pattern. As shown, the pattern selection algorithm works effectively by reducing the runtime by more than 80% comparing with the all-clips simulation, and keeps the lithography performance at 80 nm DOF at 5% EL, the same result as the all-clips set. The conventional manually-selected-clips failed to represent the full-clip set lithography performance, dropping the DOF down to 30 nm at 5%EL.

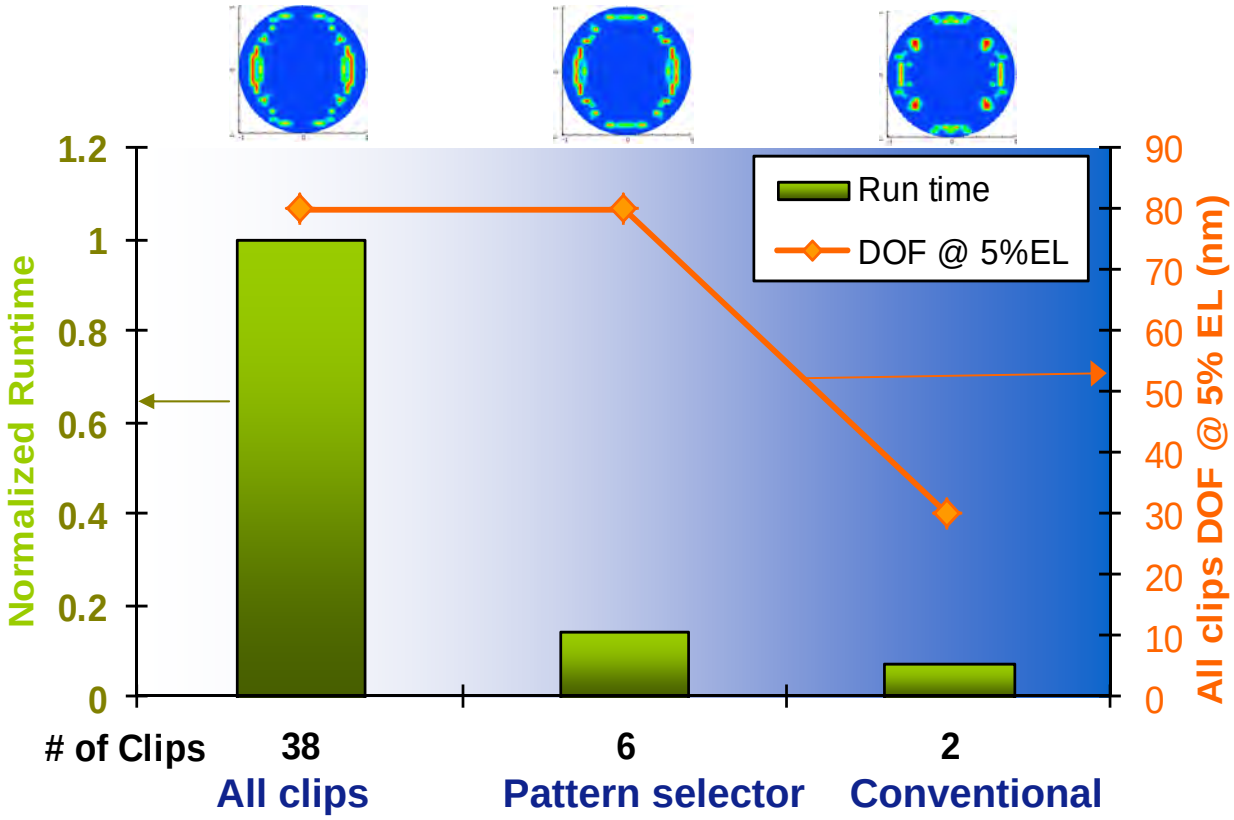

Figure 2. SMO pattern selection result of a 22 nm node.

The second case was a 3X technology node with 50 clips in-all. After pattern selections, the representative selected-clips are reduced to 12 clips. Again, the conventional manually-selected-clips include a SRAM and the densest array. Figure 3 shows results similar to the first case; the conventional manually-selected-clips failed to represent the full-clip set lithography performance, while the pattern selection works effectively.

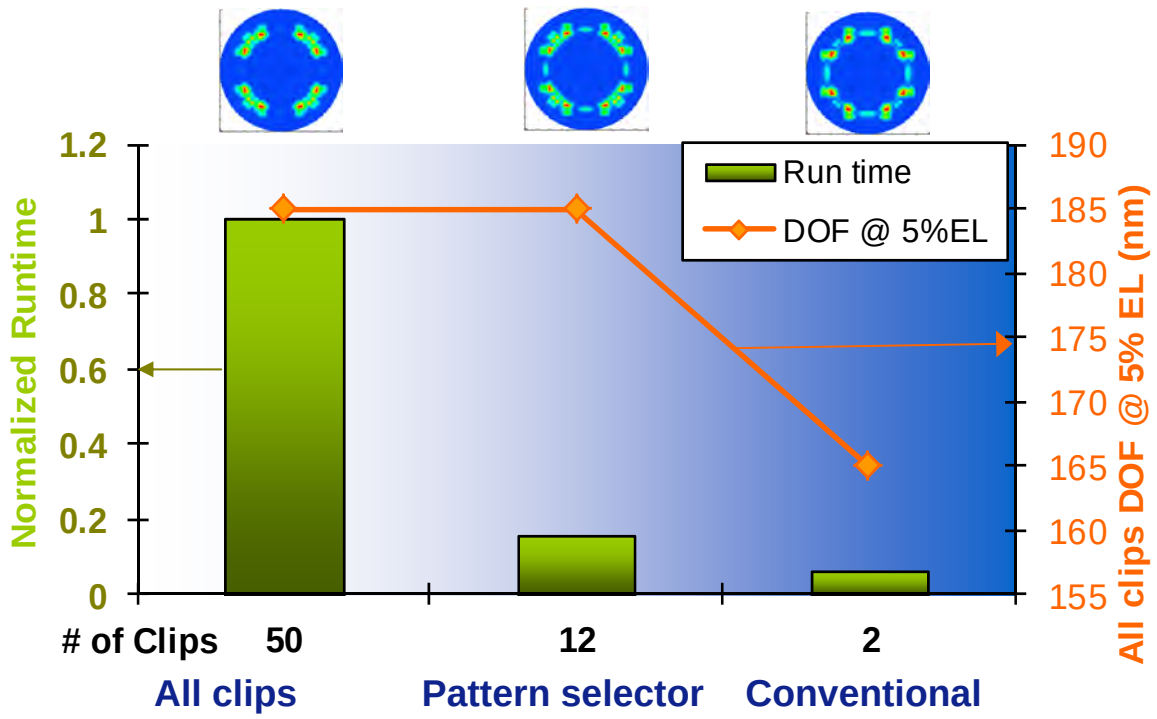

Figure 3. SMO pattern selection result of a 3X technology node.

It is noted that during the stage for optimizing SMO source for full-chip application, RET/MB-OPC recipe will also need to be developed to meet the required process window with the least mask complexity and acceptable runtime during full-chip tapeout. A cost effective full-chip mask optimization is described in the next Section.

3. FMO FOR FULL-CHIP MASK OPTIMIZATION

3.1 FMO

Traditionally, optical proximity correction (OPC) was used to achieve on-target patterning requirements. However, due to the push for low-k1 solutions, mask corrections need to accommodate additional pattern requirements, such as depth-of-focus, exposure-latitude, MEEF, etc. Also, the corrected mask has to meet mask rule constraints (MRC) and have the least mask complexity. Advanced computational lithography techniques were developed to achieve those demands. These techniques include model based sub-resolution assist features (MB-SRAF), freeform MB-SRAF, multiple variable solvers (MVS), process-window solvers, co-optimization of SRAF and main pattern methods, MEEF solvers, etc. All those techniques involve increased computational load resulting in longer tapeout time, increased mask manufacture difficulties and cost.

To mitigate runtime and cost increases and achieve required process-windows, in this paper, we proposed the flexible mask optimization (FMO) framework. Figure 4 illustrates this framework. In the Figure, LMC is Brion's RET/OPC verification tool, Tachyon LMC. The qualified SMO source for full-chip application is used for the product tapeout. In FMO, OPC1 can use a less expensive computation method. The OPC1 result is then verified by LMC1. From the LMC1

result, few critical areas or process-window limiters needing further process-window enhancements are extracted. In OPC2, these critical areas are locally corrected by a more advanced computation method, and boundary-conflicts between the different methods are resolved. Subsequently, LMC2 is applied. This process can be repeated as needed. The final result is an optimized mask that meets the required process window with acceptable runtime and mask complexity.

In the next section, we briefly review our capabilities Tachyon MB-SRAF, which is a unique model-based technique using a scatter-bar guidance-map. MB-SRAF will be used in the following section to demonstrate the FMO capability.

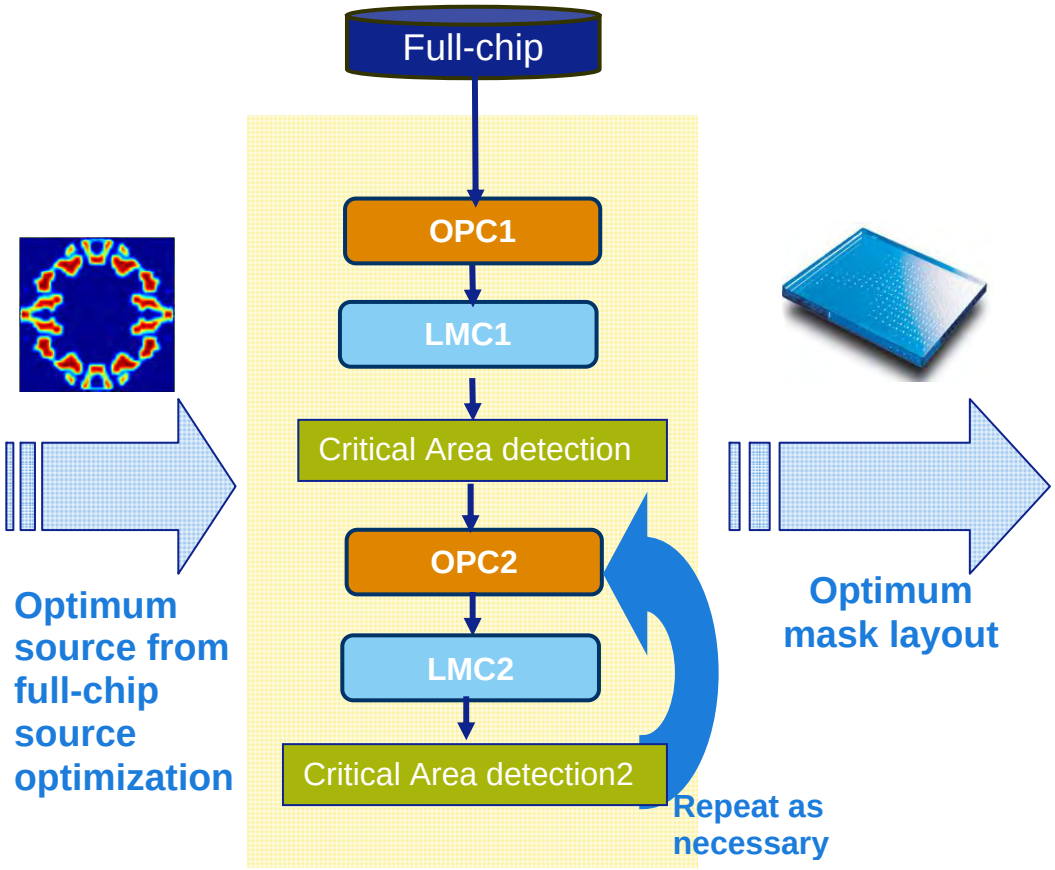

Figure 4. FMO framework.

3.2 Tachyon MB-SRAF

The Tachyon SRAF guidance-map (SGM) [6] has been widely accepted and successfully used in several manufacturing sites [7]. The map is an image of the simulation domain, in which each pixel value is calculated based on the sensitivity of improving process window on the desired pattern. SRAFs are then extracted on high sensitivity pixels and cleaned up to comply with MRC constraints. The innovative construction of the guidance map provides fast runtime. Figure 5 [6] shows the runtime comparisons of Tachyon MB-SRAF OPC with a rule-based (RB) SRAF OPC on a realistic 28 nm logic contact array with 6x6mm^2 chip-area. As shown, the runtime of Tachyon MB-SRAF OPC is about 1.5X of RB-SRAF OPC.

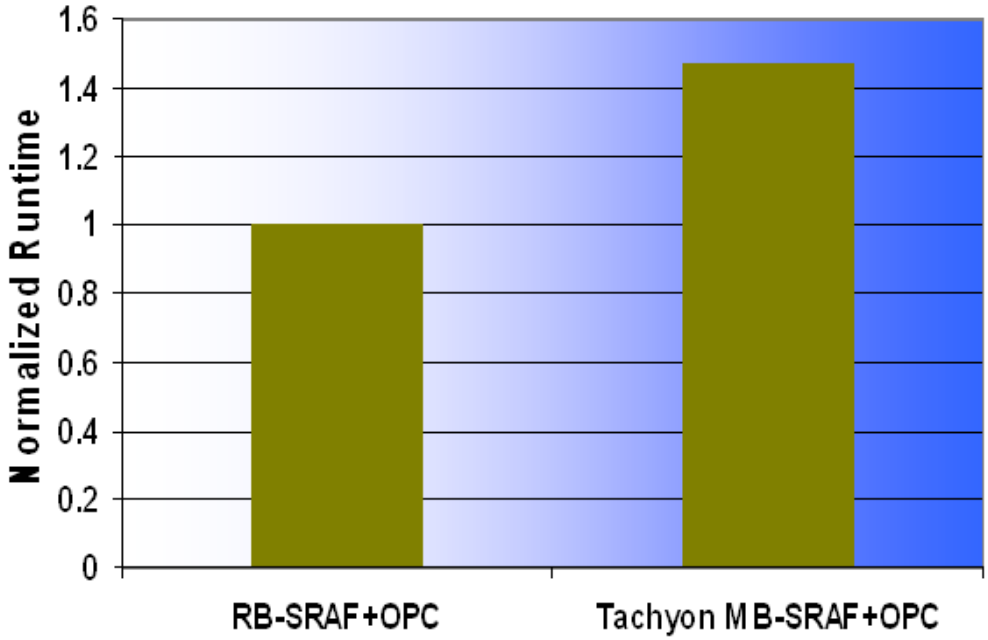

Figure 5. Tachyon MB-SRAF OPC runtime comparison with a RB-SRAF OPC.

In practice, defocus models can be used to modulate SRAF placements to achieve desired depth-of-focus (DOF) and exposure latitude (EL). Based on a realistic chip, Figure 6 shows results of exposure-defocus (ED) curves with varying defocus models used for SGM SRAF placements. In which, defocus value of A is less than B, and defocus value of B is less than C. With higher defocus values, the higher DOF is achieved at the expense of lower EL. It is noted that the value of defocus is as a tuning parameter during MB-SRAF recipe development to achieve the desired process window.

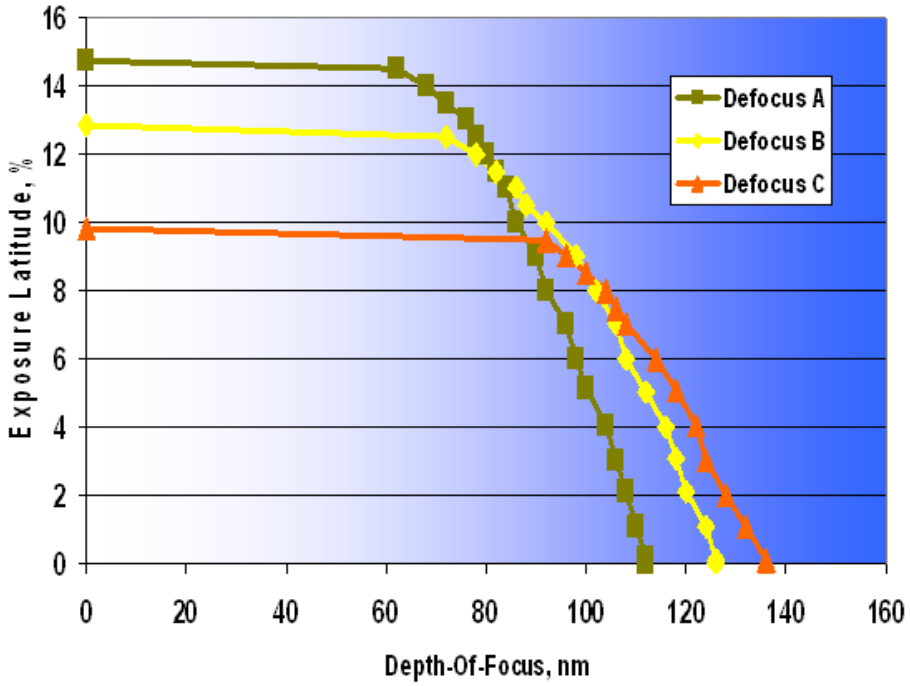

Figure 6. Results of ED curves with different defocus models used in Tachyon MB-SRAF placements.

The SGM can be used to place either rectangle SRAF (RS-SRAF) or freeform SRAF (FF-SRAF). Both FF-SRAF and RS-SRAF support full-chip capability. Figure 7 are examples of SGM produced FF-SRAF and RS-SRAF. For RS-SRAF, SGM complies with MRC constraints. Due to MRC constraints, RS-SRAF may not achieve the maximum possible process-window (PW). In practice, the trade-off between MRC and PW needs to be considered. Figure 8 shows results of the trade-off between MRC and PW on realistic design data. For this particular case, 50 nm MRC has 90 nm DOF@5%EL, while 15 nm MRC can achieve 100 nm DOF@5%EL. Tighter MRC provides a better PW, but at the expense of mask write and inspection costs.

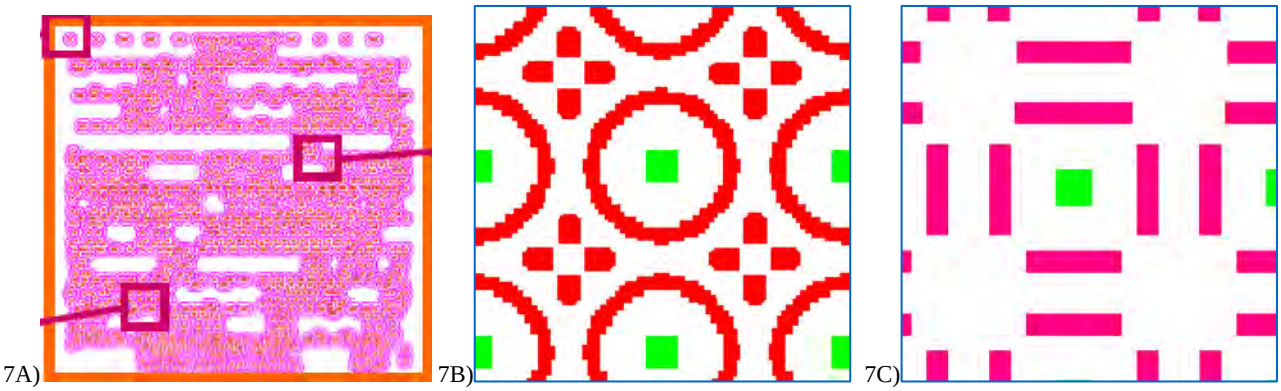

Figure 7. Examples of SGM produced Freeform and Rectangle SRAF. 7A) FF-SRAF on a full chip [7]. 7B) FF-SRAF on contact arrays. 7C) RS-SRAF on the same contact arrays.

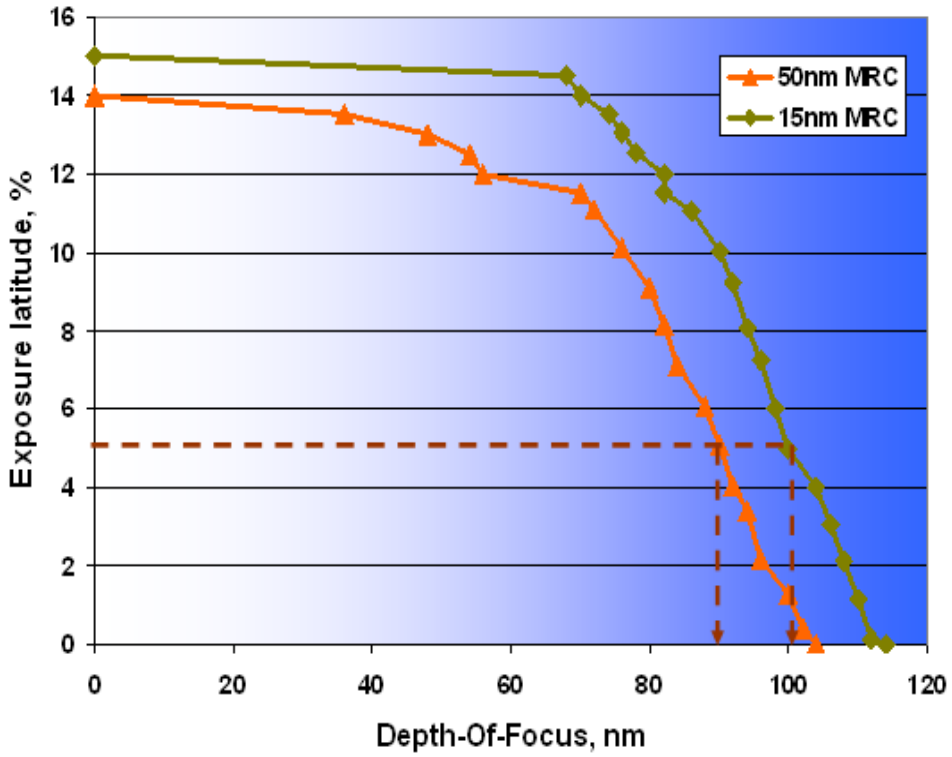

Figure 8. MRC effects on PW degradation. Tighter MRC provides better PW. In this case, 50 nm MRC has 90 nm DOF@ 5%EL, while 15 nm MRC can achieve 100 nm DOF @5%EL.

Another PW loss of RS-SRAF comes from “rectangularization”. “Rectangularization” is the process to align SRAFs to the required Manhattan rectangle shapes. During that process, some SRAF may not be able to align and are forced to be deleted. Tachyon MB-SRAF uses a model-based signal to mitigate the rectangularization-induced PW loss. FF-SRAF, which does not have this constraint, will provide the best PW. Figure 9 shows the trade-off between RS-SRAF and FF-SRAF. In that particular data, RS-SRAF has 100 nm DOF@5%EL. Using FF-SRAF can achieve 134 nm DOF@5%EL. However, for production tapeout, FF-SRAF faces several issues including longer tapeout throughput time, long mask write-time and difficult mask inspection and repair.

Those issues can be mitigated by FMO. In the next section, we use RS-SRAF and FF-SRAF as an example to demonstrate the FMO benefits.

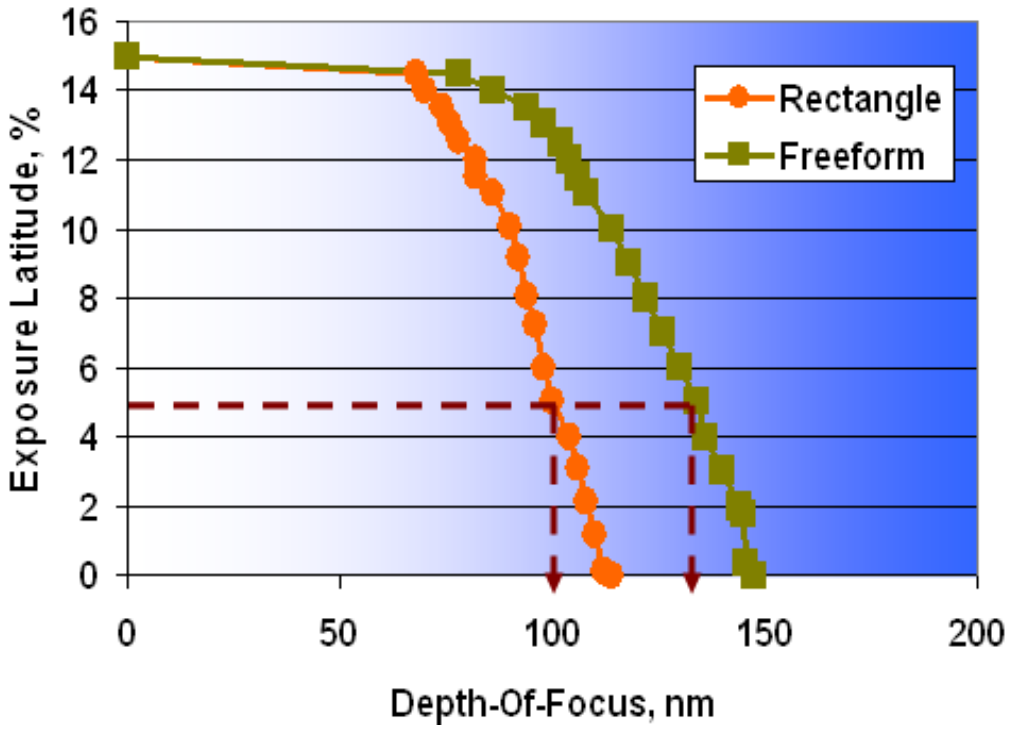

Figure 9. PW differences between Rectangle SRAF and Freeform SRAF. In this case, RS-SRAF has 100 nm DOF @ 5%EL, while FF-SRAF can achieve 134 nm DOF@5%EL.

3.3 A FMO example

In this section, we demonstrate FMO using RS-SRAF and FF-SRAF. In OPC1, RS-SRAF and OPC are applied to correct a realistic 40x40 um^2, 2X node data. As demonstrated in Figure 10, LMC1 detects the DOF limited at defocus F (-80 nm) and F (75 nm) respectively, then the critical area of those locations are extracted.

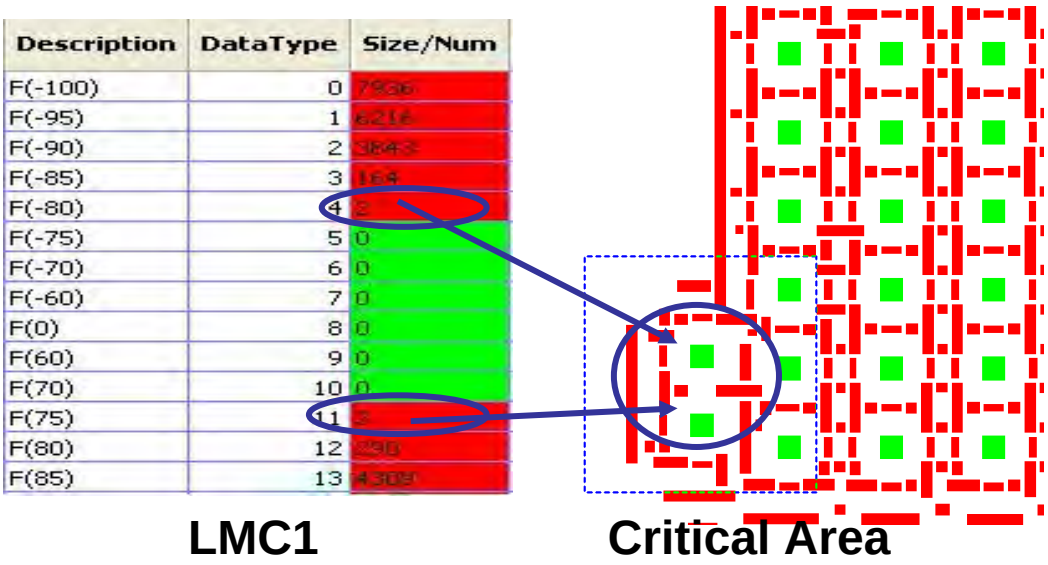

Figure 10. Demonstration of the LMC1 and critical area in FMO. RS-SRAF is generated in OPC1.

In OPC2, freeform SRAF are placed in those identified critical areas, and OPC2 corrections are applied. Around critical area boundaries, MRC conflicts between the newly placed freeform SRAF and the original rectangle SRAF are resolved based on user defined priorities. OPC corrections are applied inside and around those boundaries and weighted with the OPC1 results outside of boundaries to ensure correction quality. Figure 11 shows the DOF improvements, no defects at defocus conditions F (-80) and F (75) anymore. Figure 12 shows the ED improvements of this FMO example. OPC1 with RS-SRAF has 146 nm DOF @ 6% EL. After FMO, OPC2 with RS + FF SRAF improves DOF@6%EL to 158 nm.

In production, applying FF-SRAF where it is most needed dramatically reduces the mask complexity, electron beam write and inspection time. In addition, the critical area information can be fed to the mask inspection stage, where different inspection rules can be used in a manageable fashion. FMO's robust framework provides flexibility for users to use different methods in different stages and is not limited to the example described in this paper. Other approaches can be performed, such as using more aggressive MRC SRAF rules, multiple-variable solvers, co-optimization, process-window solvers, MEEF solvers, etc. The goal is to achieve the required lithography performance with the least cost.

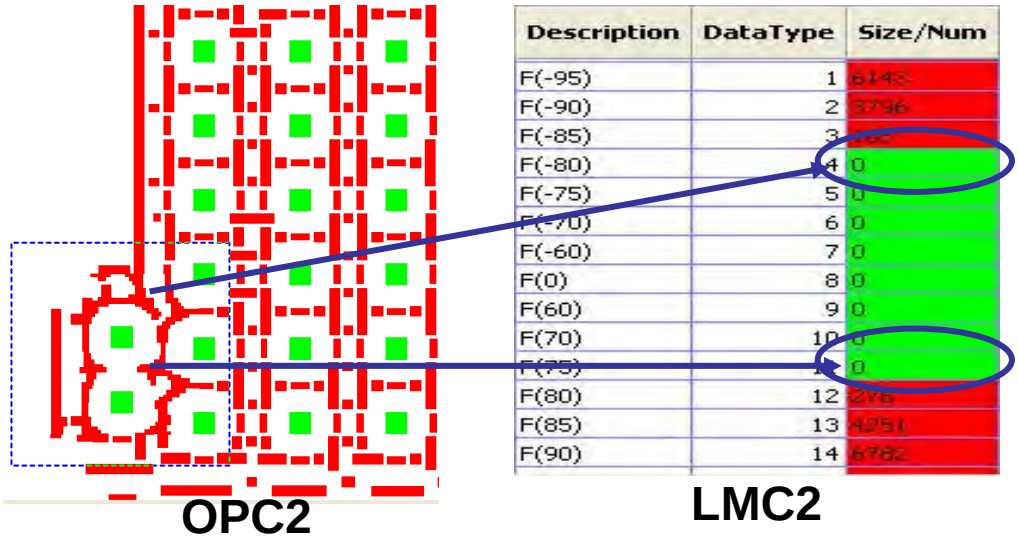

Figure 11. Freeform SRAF applied in OPC2 and LMC2 shows no more defects in defocus conditions F (-80) and F (75).

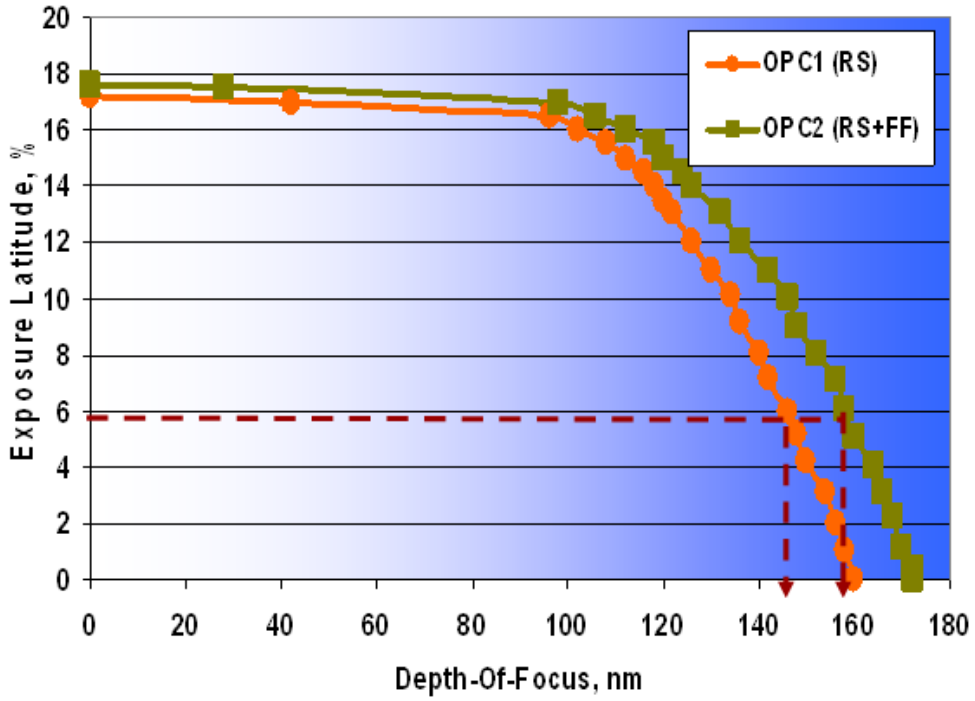

Figure 12. An example of ED improvement using FMO.

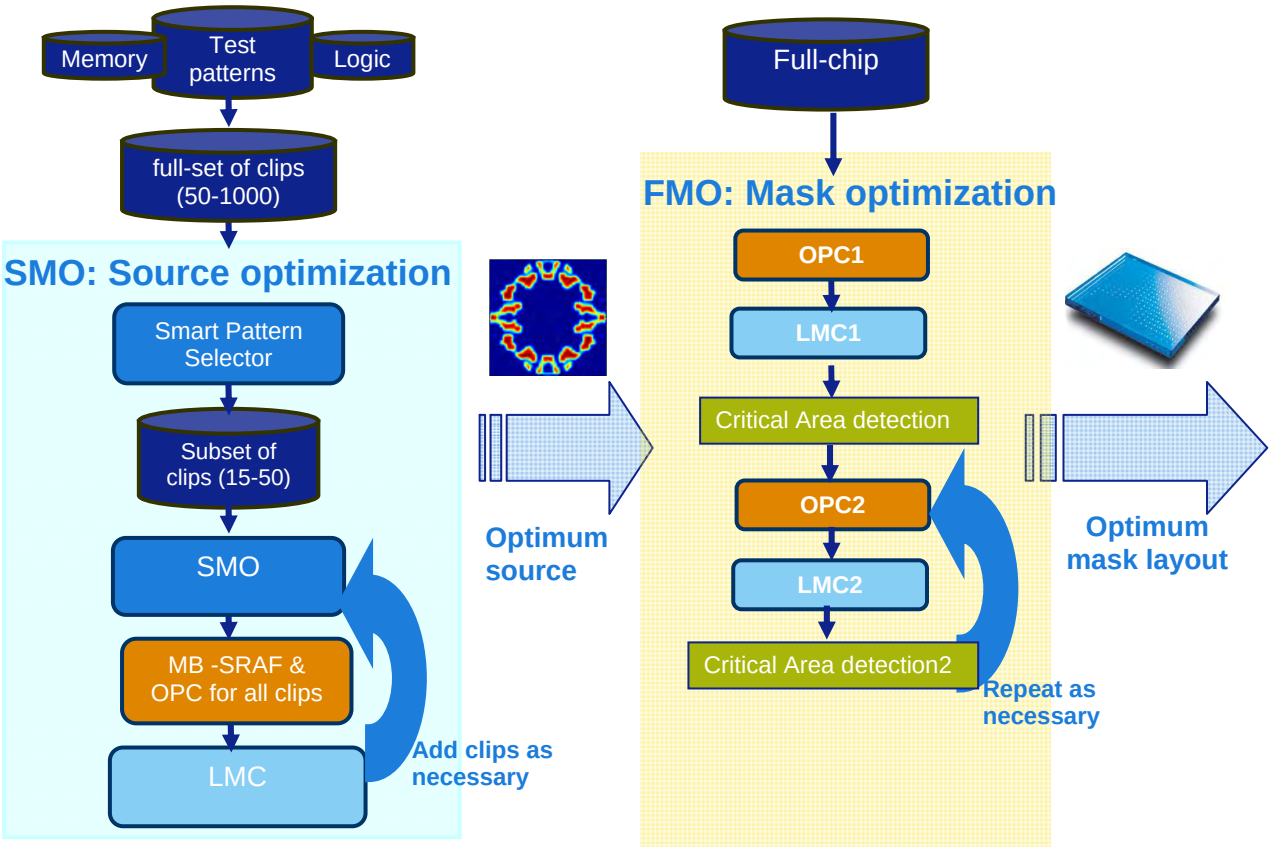

Figure 13. Complete flow of the full-chip source and mask optimization.

4. CONCLUSIONS

In this paper we have demonstrated a cost-effective full-chip source and mask optimization. For optimizing the source, we have developed an innovative pattern selection algorithm, and the method of integrating RET OPC recipe to qualify the SMO source for full-chip applications. To enable technology for low-k1 lithography, we have developed an efficient and effective MB-SRAF technique, which has been widely accepted by customers. To reduce cost in production tapeout, we have developed the FMO framework with boundary healing and management, allowing users to use advanced computation techniques only in where they are needed. Figure 13 highlights the complete flow of this full-chip source and mask optimizations.

REFERENCES

[1] Robert Socha, Xuelong Shi and David LeHoty, "Simultaneous source mask optimization (SMO)", SPIE Vol. 5853, pp 180-193 (2005).

[2] Stephen Hsu, Luoqi Chen, Zhipan Li, Sean Park, Keith Gronlund, Hua-yu Liu, Neal Callan, Robert Socha and Steve Hansen, "An Innovative Source-Mask co-Optimization (SMO) Method for Extending Low k1 Imaging", Proc. of SPIE Vol. 7140, 714010 (2008).

[3] Stephen Hsu, Zhipan Li, Luoqi Chen, Keith Gronlund, Hua-yu Liu and Robert Socha, "Source-mask co-optimization: optimize design for imaging and impact of source complexity on lithography performance", Proc. SPIE 7520, 75200D (2009).

[4] Seiji Nagahara, Kazuyuki Yoshimochi, Hiroshi Yamazaki, Kazuhiro Takeda, Takayuki Uchiyama, Stephen Hsu, Zhipan Li, Hua-yu Liu, Keith Gronlund, Terunobu Kurosawa, Jun Ye, Luoqi Chen, Hong Chen, Zheng Li, Xiaofeng Liu and Wei Liu, "SMO for 28-nm logic device and beyond: impact of source and mask complexity on lithography performance", Proc. SPIE 7640, 76401H (2010).

[5] J. Bekaert, B. Laenens, S. Verhaegen, L. Van Look, D. Trivkovic, F. Lazzarino, G. Vandenberghe, P. van Adrichem, R. Socha, S. Baron, M. C. Tsai, K. Ning, S. Hsu, H. Y. Liu, M. Mulder, A. Bouma, E. van der Heijden, O. Mouraille, K. Schreel, J. Finders, M. Dusa, J. Zimmermann, P. Gräupner, J. T. Neumann, and C. Hennerkes, "Freeform illumination sources: an experimental study of source-mask optimization for 22-nm SRAM cells", Proc. SPIE 7640, 764008 (2010).

[6] Min-Chun Tsai, Shigeki Nojima, Masahiro Miyairi, Tatsuo Nishibe, Been-Der Chen, Hanying Feng, William Wong, Zhangnan Zhu and Yen-Wen Lu , "A full chip MB-SRAF placement using the SRAF guidance map", Proc. of SPIE Vol. 7823, 78233Q (2010).

[7] Seiji Nagahara, Kazuyoshi Kawahara, Hiroshi Yamazaki, Akihiko Andoa, Masayuki Naganuma, Kazuyuki Yoshimochi, Takayuki Uchiyama, Ken Nakashima, Hidemichi Imai, Katsuya Hayano, Hidekazu Migita and Eiji Tsujimoto, "SMO mask requirements for low *k*1 lithography", Proc. of SPIE Vol. 7823, 782310 (2010).